

A complex physiology-based empirical usability evaluation method in practice

Károly Hercegi / Márton Pásztí

Received 2010-07-12

Abstract

This paper outlines the INTERFACE usability evaluation methodology developed by researchers of our department. It is based on the simultaneous assessment of Heart Period Variability (HPV), Skin Conductance (SC), and other data. One of the highlights of this methodology is its capability to identify quality attributes of software elements with a time-resolution of only a few seconds: in particular cases it can assess 2- or 3-second events. The Department of Ergonomics and Psychology at the Budapest University of Technology and Economics carried out applied research projects assessing very various software. After these, we can show different types of typical software problems identified by our method. The method of analysis allows us not only to decide what types of problems are significant to the users; however, on the other hand, the method allows us to decide, to what extent the found problems and their assessed severity concern all the users in general, or how these things depend on the type and characteristics of the users.

Keywords

Human-Computer Interaction (HCI) · usability testing and evaluation · empirical methods · case study · Heart Period Variability (HPV) · Skin Conductance (SC)

Acknowledgement

The authors would like to thank Prof. Lajos Izsó and Dr. Eszter Láng for the earlier developments, our research fellows and industrial partners for the support of the deep research, and the participants of the series of experiments for their valuable contribution.

Károly Hercegi

Department of Ergonomics and Psychology, BME, 1111 Budapest, Egry J. u. 1., Hungary
e-mail: hercegi@erg.bme.hu

Márton Pásztí

Department of Ergonomics and Psychology, BME, 1111 Budapest, Egry J. u. 1., Hungary
e-mail: tim@erg.bme.hu

1 Introduction: The background of the usability evaluation projects at the Budapest University of Technology and Economics

The main educational and research fields of the Department of Ergonomics and Psychology at the Budapest University of Technology and Economics are applied psychology and ergonomics (human factors).

The Work and Organisational Psychology research and development (R&D) projects of the Department cover the following areas:

- psychological aspects of personnel selection and socialization;
- on-the-job training and skill development;
- team-cognition;
- psychological aspects of occupational rehabilitation; etc.

The Ergonomics / Human Factors R&D projects of the Department are performed in the following areas:

- evaluation of work process, workplace and environment (typically applied for industrial and/or office environment);
- product ergonomics, human-centred product management;
- Computer Aided Anthropometric Assessment (CAAA);
- human factors of safety;
- design for all / design for special populations, assistive technologies;
- Human-Computer Interaction (HCI), including physiologically based empirical evaluation, webmining, human factors of e-learning, Computer-Supported Cooperative Work (CSCW), etc.

A significant part of the fundamental research projects of the Department of Ergonomics and Psychology have been carried out in international cooperation: i.e. with the FH Vorarlberg University of Applied Sciences, University of Maastricht, TU Berlin, University of Maryland Baltimore County.

During the last two decades, the Department of Ergonomics and Psychology gradually developed mutually useful relationships also with industrial partners, successfully accomplishing various R&D projects.

The Human-Computer Interaction R&D projects of the Department of Ergonomics and Psychology are sometimes embedded in larger work environment assessment projects, such as evaluation of office work environment covering a large scale of the human factors issues from the parameters of the physical environment through the layout of the workplaces to the software used. In other cases, stand-alone software usability evaluation projects are carried out.

To evaluate software, a wide palette of methods is used by the Department:

- traditional analytical methods, like guideline reviews, heuristic evaluation, GOMS-based analysis, cognitive walkthrough;
- interviews, focus group;
- empirical methods: recording video and logging interaction; measuring mental effort by visual critical flicker frequency (CFF), monitoring heart period variability (HPV), and in a biochemical way (measuring the cortisone level of the saliva); measuring emotional reactions by skin conductance (SC).

This paper gives an overview of a unique complex empirical method developed by the team of the Department of Ergonomics and Psychology, and shows various types of results of it.

2 Description of the INTERFACE methodology

Fig. 1 shows the conceptual arrangement of the *INTERFACE* (INTEgrated Evaluation and Research Facilities for Assessing Computer-users' Efficiency) workstation.

The advantage of the methodology applied in our practice lies in its capability of recording continuous on-line data characterizing the user's current mental effort derived from *Heart Period Variability (HPV)* and the user's emotional state indicated by *Skin Conductance (SC)* parameters simultaneously and synchronized with other characteristics of Human-Computer Interaction (HCI). This way, a very detailed picture can be obtained which serves as a reliable basis for the deeper understanding and interpretation of psychological mechanisms underlying HCI.

Elementary steps of HCI, like the different mental actions of users followed by a series of keystrokes and mouse-clicks, are the basic and usually critical components of using software. These steps can be modelled and analyzed by experts, but empirical studies of real users' interactions often highlight new HCI issues or give more objective results than expert analyses. One of the key aspects of the empirical methods is measuring *mental effort* as it is laid down e.g. in the earlier international standard of software product evaluation (ISO/IEC 9126:1991). Hence we need methods capable of monitoring users' current mental effort during these *elementary* steps.

To attain the above, a complex methodology was developed earlier at the Budapest University of Technology and Economics, by Prof. Lajos Izsó and his team (Izsó & Láng, 2000[9], Izsó, 2001[7], Izsó & Hercegi, 2004[8], Hercegi & al. 2006[4], Hercegi & al., 2009[5]). This paper presents the improved methodology and typical and new results.

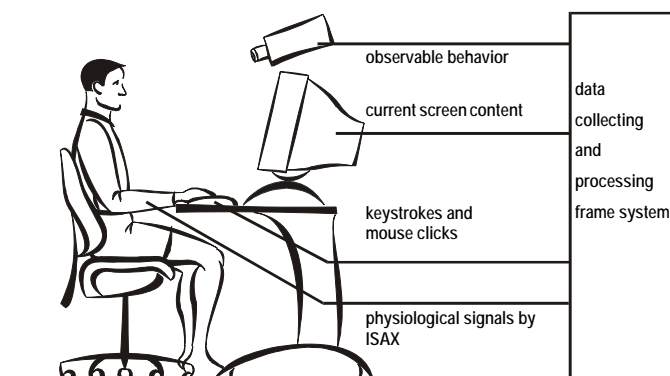


Fig. 1. Conceptual arrangement of the INTERFACE user interface testing workstation.

The INTERFACE simultaneously investigates the following:

- Users' observable actions and behaviour
 - keystroke and mouse events;
 - video record of the current screen content;
 - video records of users' behaviour: (1) mimics, (2) posture and gestures.
- Psycho-physiological parameters
 - Power spectrum of Heart Period Variability (HPV), regarded as an objective measure of current mental effort – we apply this signal successfully since more than 15 years (Izsó & Láng, 2000[9], Izsó, 2001[7], Hercegi & al., 2006[4]);
 - Skin Conductance (SC) parameters, indicating mainly the emotional reactions – recently integrated into our system (Hercegi & al., 2009[5]).

In addition to observable elements of behaviour, the applied complex method also includes traditional interviews to assess mental models, subjective feelings, and the users' opinions about their perceived task difficulty and experienced fatigue.

Recording these various data simultaneously requires a more sophisticated technical background than other empirical methods based on only personal observation or simple video recording. However, multiple channels enable researchers to concentrate on the channels that highlight the importance of various parts of the current event flow.

2.1 Fundamentals of the methodology

2.1.1 Assessing users' performance and behaviour

Performance measures are useful in general and in other projects of ours, but in the current study, we will apply them only for particular aspects of the interaction.

Recording *users' behaviour* has outstanding importance. The video recording of the user's face and activity is an extremely rich source of psychological information as it directly reflects the mental state (e.g. boredom, routine activity in familiar environment, attention-demanding task, getting lost, emotions like frustration, anger, joy, etc.). To analyze this channel, we are working on integrating a new, sophisticated method into our INTERFACE methodology.

2.1.2 Assessing mental effort via analyzing users' HPV power spectrum

Sometimes the Heart Rate (HR) itself is used in usability evaluations; however it is not a sensitive parameter in point of usability.

The deviation (or variance) of it can give us better results, but the sources of the variability include physiological mechanisms independent from the usability aspects. Because of this, further analysis of Heart Rate Variability (HRV) is needed. Although in the literature the term "Heart Rate Variability" (HRV) is more frequently mentioned, we prefer the similar expression "Heart Period Variability" (HPV), where the periods of time between the heart beats are simply the reciprocal values of the heart rates: in practice, the periods of heart beats can be analyzed more directly, and they can be more expressive.

After analyzing this variability, a number of studies (Mulder & Mulder-Hajonides van der Meulen, 1973[21], Rowe & al., 1998[26], Izsó & Láng, 2000[9], Izsó, 2001[7], Lin & Imamiya, 2006[18], Orsilia & al., 2008[24]) have shown that an increase in mental load causes a decrease in the so-called mid-frequency (MF) peak of the Heart Period Variability (HPV) power spectrum.

To assess the spectral components of HPV power spectra, an integrated system called ISAX (Integrated System for Ambulatory Cardio-respiratory data acquisition and Spectral analysis) was developed and successfully used by Dr. Eszter Láng and her team (Láng & al., 1989[13], Láng & al., 1998[12], Izsó & Láng, 2000[9]). This equipment and the related method have been integrated into our INTERFACE system.

The main advantage of our method over the previously existing HPV-based methods is that the MF component of HPV shows changes in mental effort *in the time range of several seconds* (as opposed to the earlier methods with a resolution of tens of seconds at the best). This feature was achieved by an appropriate windowing data processing technique, and application of an all-pole auto-regressive model with built-in recursive Akaike's Final Prediction Error criteria and a modified Burg's algorithm.

2.1.3 Assessing emotional responses via analyzing users' skin conductance parameters

Changes in the electrical activity of the skin (the so-called Electrodermal Activity – EDA) can be produced by various physical and emotional stimuli. We use the parameters derived

from Skin Conductance (SC) responses, especially the Alternating Current (AC) component of the SC.

In contrast with our earlier experiments applying Heart Period Variability (HPV), measuring Skin Conductance (SC) in our INTERFACE methodology is relatively new to us¹. We are working on it to complement the INTERFACE system with a component focusing mainly on the *emotional* aspects of the HCI, in addition to our well-tried approach of *mental effort*.

2.1.4 Considering applying other physiological signals

Although there are other techniques for measuring mental effort and emotions, either they are more difficult to evaluate and more invasive (e.g. the Electroencephalograph – EEG), or they give an overall, averaged indicator for a relatively long period of time, from minutes to hours (e.g. the visual critical flicker frequency (CFF) and the practical realizations of bio-chemical measures).

EEG needs more electrodes than the ECG, its electrodes have to be positioned more carefully, and the participants experience it as more invasive. Furthermore, it results more complex curves – i.e. the effects of the eye blinks have to be filtered (Luu & al., 2009[19]), and it is only one example. So, if there is a simpler method (the ECG) to objectively identify mental effort, the simpler method is preferred to apply.

However, applying EEG can be a potential direction of *further developments* of INTERFACE methodology: not to simply identify mental effort, but (1) identify more complex mental or emotional state patterns (i.e. using complex methods to analyze the complex curves (Brouwer & al., 2009[1], Lee & al., 2009[17])), or (2) attempt to localize the active brain regions (using 128- or 256-channel Dense Array EEG (dEEG) (Huang & al., 2009[6], Srinivasan & al., 2009[27])).

Electromyography (EMG) measures muscle activity by detecting surface voltages that occur when a muscle is contracted. In isometric conditions (no movement) EMG is closely correlated with muscle tension. When used on the jaw, EMG provides a very good indicator of tension in an individual due to jaw clenching. On the face, EMG has been used to distinguish between positive and negative emotions. EMG activity over the brow (frown muscle) region is lower and EMG activity over the cheek (smile muscle) is higher when emotions are mildly positive, as opposed to mildly negative (Mandryk & al., 2006[20]). Because of the small sizes (the distance between the electrodes is only about 5 millimetres) and the closeness of the muscles of the different mimic functions, the electrodes have to be positioned extremely carefully (Park, 2009[25]). Furthermore, the participants experience the electrodes on the face or head as

¹ An interesting series of experiments using the new version of the ISAX to analyze SC responses is finished by one of our colleagues (Laufer, 2007[14], Laufer & Németh, 2008[15], Laufer & Németh, 2008[16]). It is a good example of the promising way to use data mining techniques in empirical usability studies. In that case, the tool was not yet integrated into the complex INTERFACE system.

more invasive than the electrodes on the fingers measuring Skin Conductance (SC). So, if there is a simpler method (measuring SC) to identify emotional reactions, the simpler method is preferred to apply (in spite of the EMG's potential capability of differentiate positive and negative emotions).

Measuring mental effort by visual critical flicker frequency (CFF), and in a biochemical way (measuring the cortisone level of the saliva) – as they were mentioned in the introduction – are applied by our Department (Izsó, 2001[7]). However, these methods give only an overall, averaged indicator for a relatively long period of time, from minutes to hours – this is not the time resolution that is goaled by the INTERFACE methodology.

Eye-tracking is a promising direction of *further developments* of INTERFACE methodology: (1) it is reliably capable of localizing the user interface elements that cause high mental effort or emotional reactions identified by the other physiological channels, and (2) it can be analyzed deeper, deriving parameters referring the state of the nervous system (Obinata & al., 2009[22]).

Applying *pupillometry* (measuring the diameter of the pupil) is a measurement option that is often included in the capabilities of eye-tracker equipments. It is sensitive to both the mental effort and the emotions (Oliveira & al., 2009[23], Tullis & Albert, 2008[28]). It can be capable of validate the other physiological channels of the INTERFACE methodology.

Eye-tracking and pupillometry are used in our *ongoing INTERFACE research*: the concept of the research have already been published (Komlodi & Hercegi, 2009[10], Komlodi & Hercegi, 2009[11]), the results will be published in the next years.

2.2 Experimental arrangement

Fig. 2 shows a typical experimental arrangement of the INTERFACE methodology.

This arrangement was applied during the sessions of one of our recent researches. The Department of Ergonomics and Psychology at the Budapest University of Technology and Economics latched on to the ongoing software development process of the Generali-Providencia Insurance Co. to carry out the usability assessment of their customer service software (Hercegi & al., 2009[5]). In this case, the series of experiments was carried out on-site, in the real call centre.

The chosen workstation was located in the corner of the operators' room, in order not to disturb others. The experimenter sat next to the participant. Behind them, the team leader's glass wall could be found, and so our staff could sit and observe the sessions and make simulated phone calls from behind this pane.

Three Electrocardiogram (ECG) electrodes were put on the users' torsos and there were two electrodes placed on their left hands (in case of right-handed persons) for measuring Skin Conductance (SC). The arrangement of the video cameras and the other equipments can be seen in the figure.



Fig. 2. The experimental arrangement applied during the sessions of the INTERFACE usability test, installed on a standard workstation of a call centre.

2.3 The viewer screen of the INTERFACE software

The main point is that the INTERFACE Viewer software can play the records of the different data channels simultaneously. Fig. 3 shows the INTERFACE Viewer screen with a record of the empirical test of the above mentioned call centre software. This figure also shows the typical pattern of mental effort observable both on the HPV curve and in the video images.

2.4 Experiences

During the years, the Department of Ergonomics and Psychology could gain experiences by the INTERFACE methodology applying it to assess very various types of software:

- electronic mailing systems of the *Dutch post* company (PTT) (Izsó, 2001[7]);
- simulation centre of the *Paks Nuclear Power Plant* (Izsó, 2001[7]);
- Computerized Directory Assistance Services software of the *MATÁV Hungarian telecom* company (Izsó & Láng, 2000[9], Izsó, 2001[7]);
- an educational *multimedia* CD (Izsó, 2001[7]);
- hypertext-based *e-learning* titles developed by us (Hercegi & al., 2006[4], Hercegi & Kiss, 2009[3]);
- *ArchiCAD*, the market leading architectural Computer Aided Design (CAD) software released by the *Graphisoft*;
- a simple computer *game*;
- WAP-based software of *Nokia*;
- *air traffic control* system of the *Eurocontrol*;

Fig. 4. Examples of very various types of software assessed by the INTERFACE methodology. The selected moment of each record shows similar situation to that could be seen in Fig. 3: in each case, currently the user makes significant mental effort, as it is shown by the facial expression, gesture, and posture, and the low value of the last green profile curve of the Mid-Frequency (MF) power of the Heart Period Variability (HPV) at the cross-hair.

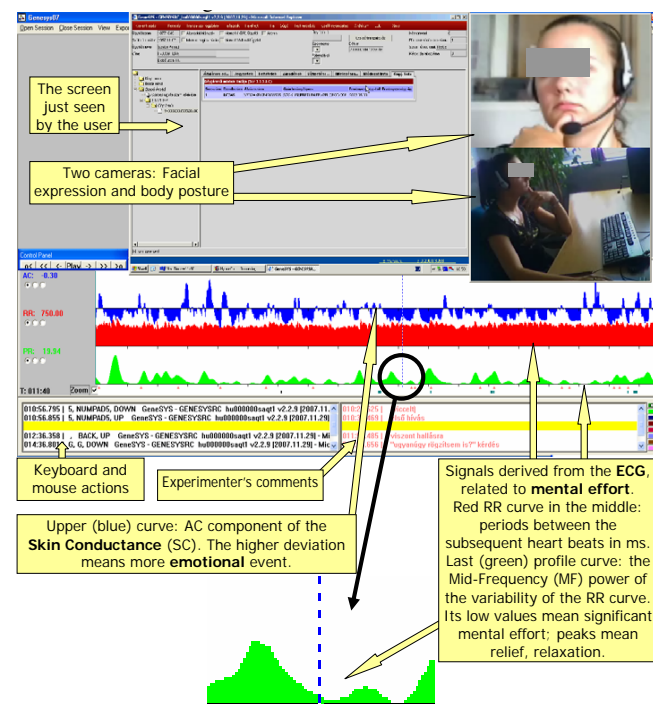


Fig. 3. The INTERFACE Viewer screen with a record of the empirical test of a call centre software. As it can clearly be seen, currently the user makes significant mental effort – it is shown by the facial expression and gesture, and the low value of the last profile curve of the Mid-Frequency (MF) power of the Heart Period Variability (HPV) at the cross-hair. (The currently displayed curves in the window show the history of 20 minutes. The enlarged valley of the profile curve shows a period of 38 seconds. It is a robust mental effort needed period, selected as illustration – however the much smaller valleys and peaks can be analyzed and interpreted as well.)

- *flight simulator* software (and various hardware) developed at our university;
- the *web-based* editor interface of the *Moodle* Learning Management System (LMS);
- Work and Force Management Software of the *Magyar Telekom*;
- customer centre software developed and used by the *Generali-Providencia Insurance Co.* (Hercegi & al., 2009[5]).

Fig. 4 shows examples for the diversity of the software assessed.

3 Validation

Validation results of earlier researches have already been published (Izsó & Láng, 2000[9], Izsó, 2001[7]). This paper aims to give an overview of the INTERFACE methodology, so some recent validation results follows.

At the beginning of each session of our series of experiments, we used to perform a “calibration” phase.

First, the user is asked to relax for cc. two minutes.

The relaxation is followed by two-minute mental effort: mental arithmetic. The result of the counting is not too important; the only goal is generating mental effort.

The curves shown in the upper and lower parts of the Fig. 5 were recorded during two different sessions (No. 11, No. 10) of the mentioned series of experiments carried out in the call centre of the insurance company.

In both cases, the three curves are the curves of the AC of Skin Conductance (SC), the RR curves (Heart Periods), and the

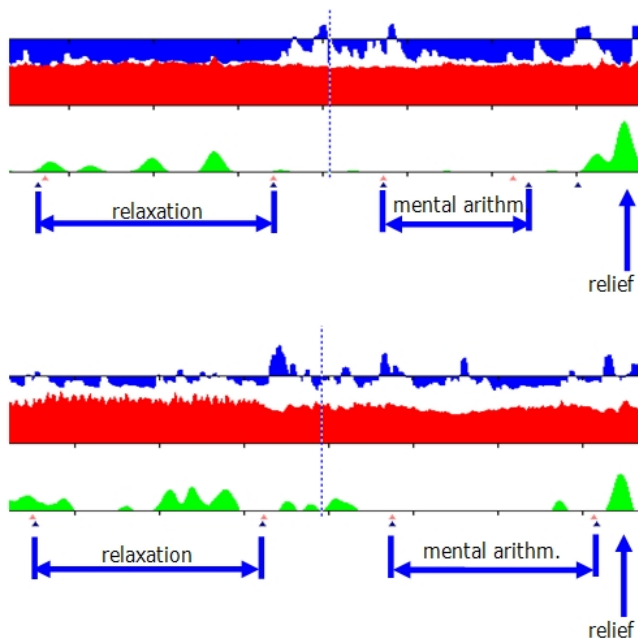


Fig. 5. The typical pattern of the relaxation and mental arithmetics in cases of two participants. (In these cases the relaxations were 2 min 53 sec and 2 min 41 sec, the mental arithmetic periods were 1 min 42 sec and 2 min 24 sec.

profile curves of the Mid-Frequency (MF) power of Heart Period Variability (HPV).

The curve of AC of SC is relatively smooth during both the relaxation and the mental arithmetic. During these sections, there are not any emotional peaks, and these two participants can be characterized as “stable” type according to the typology of physiology. However, the beginnings and the ends of the sections are followed by peaks.

During relaxation, the MF component of the HPV increases, so the RR curve has zigzags, and the profile curve is relatively high. (In case of perfect relaxation, the profile curve should be consistently high. However, this is not expected in this experimental situation: due to the real life situation previously described, the participants were disturbed by their colleagues’ calls and talks; and, of course, they were just wired and begun being observed.) The curve can be considered as high, especially in comparison with the next section.

During mental arithmetic, the RR curve gets smoother, and the profile curve is significantly low.

After the “calibration” tasks, the participants really relieve. During this short period of relief, the participants get more relaxed than during the conscious, intended relaxation: the MF of HPV profile curves have their highest peaks here.

These “calibration” tasks prove a validation of the method. The values of the MF power of the HPV were significantly higher during relaxation than during mental arithmetic. A non-parametric statistical method, the Wilcoxon Signed Ranks Test proves the difference (sig. 0.037 – Fig. 6). It is a significant difference, in spite of the non-perfect relaxation.

However, the mental arithmetic task works better: the sig-

nificance of the difference between the values of MF power of HPV during mental arithmetic and in general, during the whole software usage section is better: the Wilcoxon test results sig. 0.002.

The values of the deviation of the AC component of Skin Conductance (SC) do not differ during the relaxation and the mental arithmetic significantly. As it was described earlier, this is the expected result.

However, the deviations of the AC of SC during the relaxation and the mental arithmetic are significantly lower than in general, during the whole software usage section: the Wilcoxon tests result sig. 0.009 and sig. 0.017.

From these results, it can be stated that low values of the MF power of HPV really indicate mental efforts, and high deviations of AC of SC probably mean higher emotions. Then, in the section of software usage, moments with relatively high (and unwanted) mental efforts and high (and unwanted and not positive) emotions are looked for. This method gives the key to find the problems of the user interface.

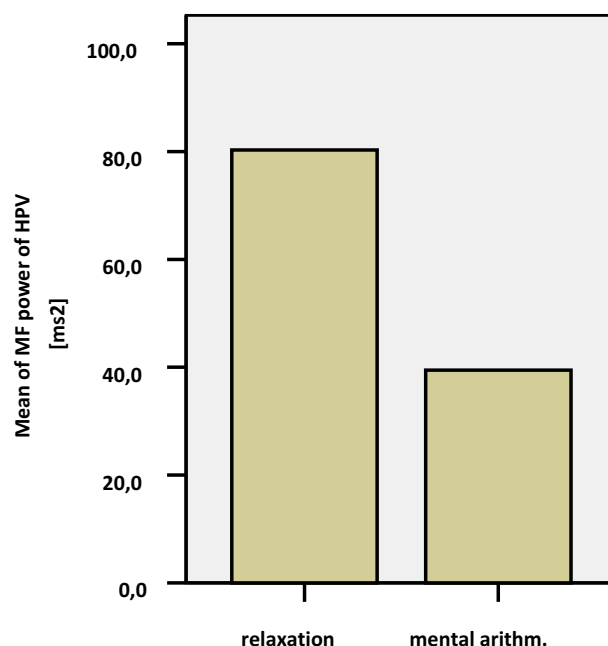


Fig. 6. Validation of measuring Mid-Frequency (MF) power of Heart Period Variability (HPV) as an indicator of mental effort: the MF power of HPV was significantly higher during relaxation than during mental arithmetic (sig.0.037).

4 Typical user interface problems identified by the INTERFACE methodology

From the series of our researches, now we highlight only two examples to show very different types of problems can be found by the INTERFACE methodology.

The following cases are from a hypertext-based multimedia e-learning development project led by us. Details of the project can be found in our other publications (Hercegi & al., 2006[4], Hercegi & Kiss, 2009[3]).

The participants of the study were 21 students of two voca-

tional secondary schools. They were performing learning sessions fitting to their studies with the help of the multimedia title “Basics of Information Technology” (in Hungarian).

Because we were interested in the effects of the individual differences, we collected demographic data, academic records, data about the familiarity with the computer and the Internet, and we applied a psychological test (MBTI – Myers-Briggs Type Indicator) to identify the cognitive style of the users.

The method of analysis allows us to decide what types of problems are significant to the users, and what types of problems set back the users only slightly. On the other hand, the method allows us to decide, to what extent the found problems and their assessed severity concern all the users in general, or how these things depend on the type and characteristics of the users.

We were able to focus to the quality attributes of software elements with a time-resolution of only a few seconds.

We were able to analyse the correlations of the rich data set gained by the INTERFACE system (e.g. time data, number of clicks, recorded tracking data in the hypermedia space, etc.) together with data obtained from the questionnaires and psychological tests mentioned above.

4.1 A usability problem which we originally intended to focus on; that affects all type of users

Fig. 7 shows a simple, but not rare problem: here the images were logically expected to be hot links, but they were not.

It had been a known problem before we started the series of experiments. However, we were interested in measuring the severity of it.

It is a very simple problem, so neither of the participants complained about it. However, this inconsistency objectively resulted unnecessary loads on the users and a substantial waste of time. And, as it can be shown, that affects all type of users.

(Note: Also the advantage of the high temporal resolution can clearly be seen in Fig. 7: the fine time structure of the HPV profile during the three subsequent clicks gives well established basis for interpretation.)

71% of the participants clicked on the images first (ineffectively). They found the real hot links (the text instead of the images) after an average of 5.3 seconds’ waste of time. The maximum delay was 80.5 sec (it is the case of Fig. 7).

The number of unnecessary clicks and the waste of time caused by this particular usability problem *do not correlate* with almost any of the other variables. (Spearman rho values have been calculated). The Kruskal-Wallis and Mann-Whitney statistical tests have not given any significant result. From practical point of view, this means that *this type of usability problem is a general problem; it affects all types of users.*

4.2 An unexpected usability problem. Its severity depends on the type and characteristics of users

The next example was unexpected for us.

In the Fig. 8, the difficulty of finding the scroll bar is demonstrated. The user discovered the scroll bar only after a helpless seven-minute trial-and-error searching. The origin of the problem was the following: the first part of the long, scrollable page – using this screen resolution – looks like a complete page: the figure and its caption are at the bottom of the current screen, as it can be seen in the upper screenshot of Fig. 8.

The average waste of time caused by this problem was 69 seconds, with the maximum of 253 seconds.

But it is not these mere data that are most interesting. One third of the users did not have any problems here: they clicked on the scroll bar 1 to 3 seconds after they had arrived to this page. However, the other two thirds of the users needed 14 to 253 seconds. For example, the girl in Fig. 8 had no less intellectual capacity or experiences with the Internet than the others. Why does this screen still represent a problem for her and the two thirds of the users, and why not for the others? *How does the severity of this usability problem depend on the type and characteristics of the users?*

The Mann-Whitney statistical tests show the following:

- The waste of time caused by this usability problem depends on the type of the school of the students: the students of the vocational school of economics wasted significantly more time here ($p=0.006$). However, in our case, all economics students were girls and most of the technical students were boys, so this effect cannot be separated from the effect of the gender: the significance level of the dependence on the gender is $p=0.031$.
- The users who read literature regularly wasted more time than the others, with the significance level of $p=0.021$.
- The users who read IT books and/or magazines regularly wasted less time than the others, with the significance level of $p=0.013$.

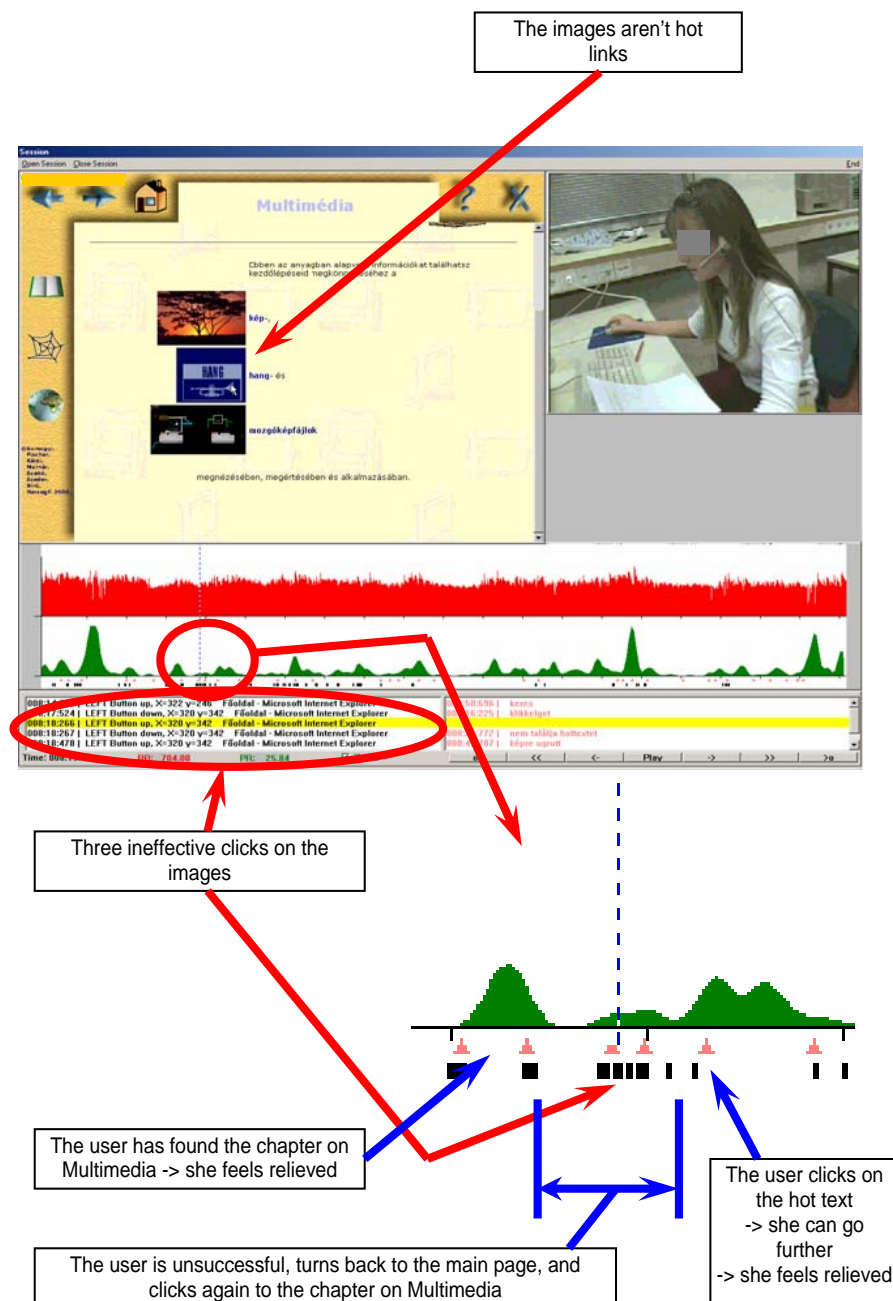
The calculations of correlations – among others – show the following:

- The students with better grade in maths wasted less time than the others; Spearman’s rho is 0,441, with the significance level $p=0.034$.
- ***The waste of time caused by this usability problem correlates strongly with scores on the Thinking–Feeling (T–F) dimension of the MBTI psychological test;*** Spearman’s rho is 0,533, with significance level of $p=0.046$. This result means, that **“thinking”-type users understood the logic** of the content and the user interface almost immediately, independently of the deceitful view of the screen; but **the users with “feeling” cognitive style had been fooled by the apparent intactness of the layout of the particular page.**

5 Conclusion

Based on the results presented here as well as in related papers, it can be stated that the INTERFACE methodology in its

Fig. 7. The user clicks on the images three times ineffectively, which resulted in a short period of unnecessary mental effort and losing several seconds.

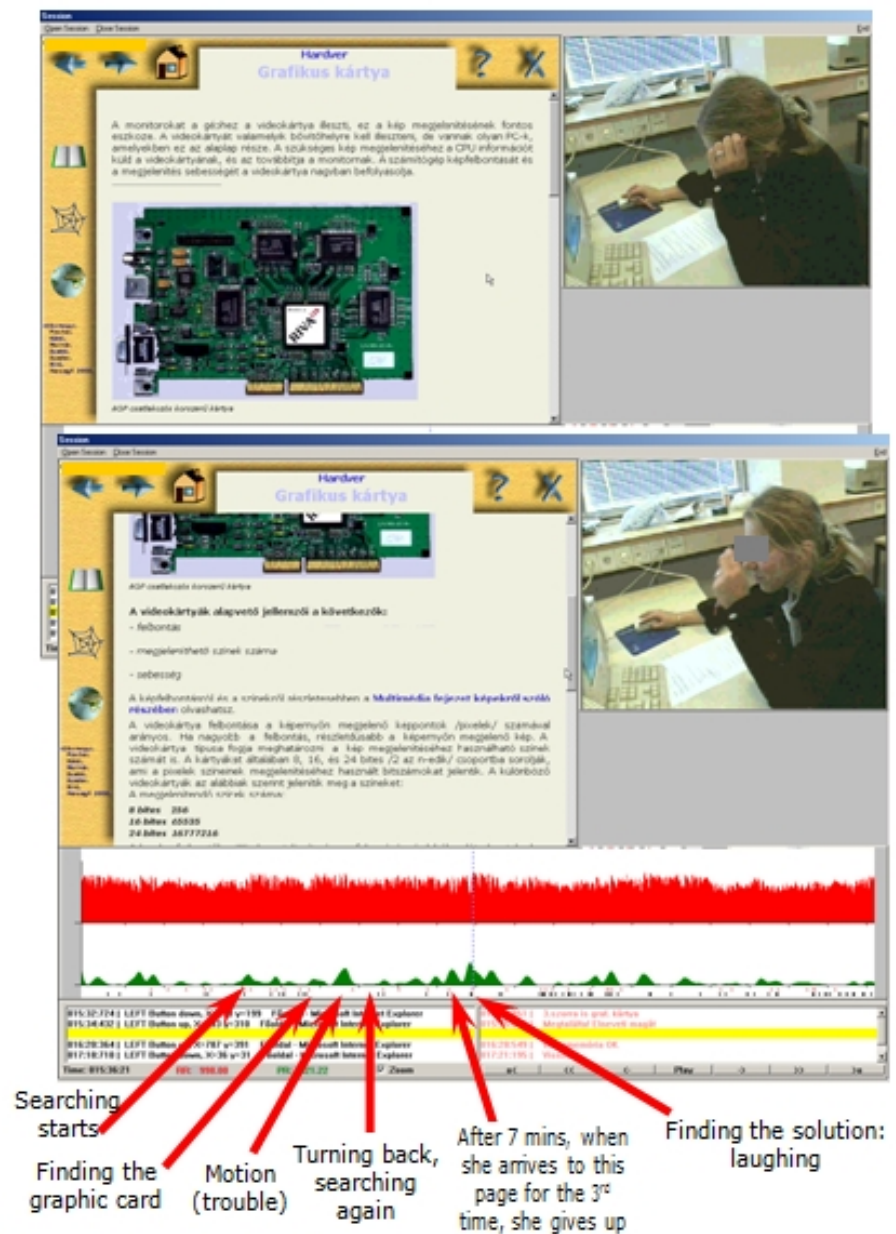


present form is capable of identifying the relative weak points of the HCI. By this methodology, it was possible to study events occurring during the HCI in such detail and objectivity that would not have been possible using other methods presently known to us. The sophisticated Heart Period Variance (HPV) profile function integrated into the INTERFACE system is a powerful tool for monitoring events in such a narrow time frame that it can practically be considered as a time-continuous recording of relevant elementary events. Measuring the Skin Conductance (SC) is a new opportunity to modulate the results.

References

- 1 Brouwer A-M, Hogervorst M, Herman P, Kooi F, *Are You Really Looking? Finding the Answer through Fixation Patterns and EEG*, 13th International Conference on Human-Computer Interaction (2009), Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience. (Schmorrow D, ed.), Vol. 5638, Springer, 2009, 329–338, DOI 10.1007/978-3-642-02812-0_39, (to appear in print).
- 2 Chen D, Vertegaal R, *Using Mental Load for Managing Interruptions in Physiologically Attentive User Interfaces*, CHI '04 (Vienna, Austria, 2004), Conference on Human Factors in Computing Systems, ACM Press, 2004, 1513–1516, DOI 10.1145/985921.986103, (to appear in print).
- 3 Hercegi K, Kiss O, *Assessment of E-Learning Material with the INTERFACE System*, Distance and E-learning in Transition (Bernath U, Szűcs A, Tait A, Vidal M, eds.), John Wiley & Sons, Hoboken, NJ, USA., 2009, 645–647.
- 4 Hercegi K, Kiss O, Bali K, Izsó L, *INTERFACE: Assessment of Human-Computer Interaction by Monitoring Physiological and Other Data with a Time-Resolution of Only a Few Seconds*, XIV. ECIS (Göteborg, Sweden, 2006), European Conference on Information Systems, ECIS Standing Committee, 2006, 2288–2299.
- 5 Hercegi K, Pászti M, Tóvölgyi S, Izsó L, *Usability Evaluation by Monitoring Physiological and Other Data Simultaneously with a time-resolution*

Fig. 8. After a 7-minute ordeal, the user gives it up, but immediately after that, she finally discovers the solution (the scroll bar) and laughs. The upper screen shows a moment when the user is in the state of giving up, while the lower screen presents the situation a bit later when she just found the scroll bar.



of only a few seconds, HCII (San Diego, California, USA, 2009), Human-Computer Interaction International, LCNS, vol. 5610, Springer, 2009, 59–68, DOI 10.1007/978-3-642-02574-7_7, (to appear in print).

6 Huang R-S, Jung T-P, Makeing S, *Spectra during Simulated Driving*, HCII (San Diego, California, USA, 2009), Augmented Cognition (Schmorrow D, ed.), LNAI, vol. 5638, Springer, 2009, 394–403, DOI 10.1007/978-3-642-02812-0_47, (to appear in print).

7 Izsó L, *Developing Evaluation Methodologies for Human-computer Interaction*, Delft University Press, Delft, The Netherlands, 2001.

8 Izsó L, Herczegfi K, *HCI Group of the Department of Ergonomics and Psychology at the Budapest University of Technology and Economics*, Ext. Abstracts CHI 2004, ACM Press, 2004, 1077–1078.

9 Izsó L, Láng E, *Heart Period Variability as Mental Effort Monitor in Human Computer Interaction*, Behaviour & Information Technology, 19(4), (2000), 297–306, DOI 10.1080/01449290050086408.

10 Komlodi A, Herczegfi K, *Cultural Differences in Information Behavior*, Third Workshop on Human-Computer Interaction and Information Retrieval, Proceedings, 2009, 78–81.

11 ———, *Exploring Cultural Differences in Information Behavior Applying Psychophysiological Methods*, Conference on Human Factors in Computing

Systems (Atlanta, Georgia, USA, 2010), Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, ACM Press, 2010, 4153–4158, DOI 10.1145/1753846.1754118, (to appear in print).

12 Láng E, Caminal P, Horváth G, Jané R, Vallverdu M, Slezsák I, Bayés de Luna A, *Spectral analysis of heart period variance (HPV) – a tool to stratify risk following myocardial infarction*, Journal of Medical Engineering and Technology, 22(6), (1998), 248–256, DOI 10.3109/03091909809010007.

13 Láng E, Szilágyi N, Métneki J, Czeizel E, 'Adám G, *A psychophysiological method for discrimination of MZ and DZ twins: A study based on impedance cardiography*, Proceedings of the 6th International Congress on Twin Studies (Rome, 1989), 1989.

14 Laufer L, *Jump is on the Skin Deep: Predicting User Behavior from Skin Conductance Level*, Ext. Abstracts HCII 2007, Springer, 2007.

15 Laufer L, Németh B, *Anticipation of Stress as an Indicator of User Interaction*, CHI2008 (Florence, Italy, 2008), Proc. Workshop on Measuring Affect in HCI: Going Beyond the Individual, ACM, 2008.

16 ———, *Predicting User Action from Skin Conductance*, IUI (Gran Canaria, Canary Islands, Spain, 2008), Proceedings of the 13th international

- conference on Intelligent user interfaces, ACM Press, 2008, 357–360, DOI 10.1145/1378773.1378829, (to appear in print).
- 17 **Lee H, Lee J, Seo S**, *Brain Response to Good and Bad Design*, Human-Computer Interaction. New Trends, LNCS, vol. 5610, Springer, 2009, 111–120.
 - 18 **Lin T, Imamiya A**, *Evaluating Usability Based on Multimodal Information: An Empirical Study*, ICMI (Banff, Canada, 2006), Proceedings of the 8th international conference on Multimodal interfaces, ACM press, 2006, 364–371, DOI 10.1145/1180995.1181063, (to appear in print).
 - 19 **Luu P, Frank R, Kerick S, Tucker D**, *Directed Components Analysis: An Analytic Method for the Removal of Biophysical Artifacts from EEG Data*, HCII (San Diego, California, USA, 2009), Foundations of Augmented Cognition Neuroergonomics and Operational Neuroscience, LNAI, vol. 5638, Springer, 2009, 411–416, DOI 10.1007/978-3-642-02812-0_49, (to appear in print).
 - 20 **Mandryk R, Atkins M, Inkpen K**, *A Continuous and Objective Evaluation of Emotional Experience with Interactive Play Environments*, CHI (Quebec, Canada, 2006), Proceedings of the SIGCHI conference on Human Factors in computing systems, ACM Press, 2006, 1027–1036, DOI 10.1145/1124772.1124926, (to appear in print).
 - 21 **Mulder G, Mulder-Hajonides van der Meulen W**, *Mental Load and the Measurement of Heart Rate Variability*, Ergonomics, **16**, (1973), 69–83, DOI 10.1080/00140137308924483.
 - 22 **Obinata G, Tokuda S, Fukuda K, Hamada H**, *Quantitative Evaluation of Mental Workload by Using Model of Involuntary Eye Movement*, HCII (San Diego, California, USA, 2009), Engineering Psychology and Cognitive Ergonomics (Harris D, ed.), LNAI, vol. 5639, Springer, 2009, 223–232, DOI 10.1007/978-3-642-02728-4_24, (to appear in print).
 - 23 **Oliveira F, Aula A, Russel D**, *Discriminating the Relevance of Web Search Results with Measures of Pupil Size*, CHI (Boston, Massachusetts, USA, 2009), Proceedings of the 27th international conference on Human factors in computing systems, ACM Press, 2009, 2209–2212, DOI 10.1145/1518701.1519038, (to appear in print).
 - 24 **Orsilia R, Virtanen M, Luukkaala T, Tarvainen M, Karjalainen P, Viik J, Savinainen M, Nygard C-H**, *Perceived Mental Stress and Reactions in Heart Rate Variability – A Pilot Study Among Employees of an Electronics Company*, International Journal of Occupational Safety and Ergonomics (JOSE), **14**(3), (2008), 275–283.
 - 25 **Park B**, *Psychophysiology as a Tool for HCI Research: Promises and Pitfalls*, HCII (San Diego, California, USA, 2009), Human-Computer Interaction. New Trends, LNCS, vol. 5610, Springer, 2009, 141–148, DOI 10.1007/978-3-642-02574-7, (to appear in print).
 - 26 **Rowe D, Sibert J, Irwin D**, *Heart Rate Variability: Indicator User State as an Aid to Human-Computer Interaction*, CHI (Los Angeles, California, USA, 1998), Proceedings of the SIGCHI conference on Human factors in computing system, ACM Press, 1998, 18–23, DOI 10.1145/274644.274709, (to appear in print).
 - 27 **Srinivasan R, Thorpe S, Deng S, Lappas T, D’Zamora M**, *Decoding Attentional Orientation from EEG Spectra*, HCII (San Diego, California, USA, 2009), Human-Computer Interaction. New Trends, LNCS, vol. 5610, Springer, 2009, 176–183, DOI 10.1007/978-3-642-02574-7_20, (to appear in print).
 - 28 **Tullis T, Albert B**, *Measuring the user Experience: Collecting, Analyzing, and Presenting Usability Metrics*, Elsevier, Burlington, MA, USA, 2008.